Maximizing Revenue for Taxi Cab Drivers through Payment Type Analysis Using Python

Submitted by

Patankar Akshay Avinash
M.Sc(23MA40019)
Department of Mathematics
Indian Institute of Technology Kharagpur

Under the guidence of

Prof. Pawan Kumar
Department of Mathematics
Indian Institute of Technology Kharagpur

Master Thesis Project 2024

Student Signature: **Supervisor Signature:**

Contents

Acknowledgements

I am heartily thankful to my supervisor **Prof. Pawan Kumar** whose encouragement, guidance, and support enabled me to develop a deeper understanding of this topic. His insightful suggestions and constant motivation have been invaluable throughout the course of this project. I sincerely appreciate his patience and the time he devoted to addressing my queries, which helped me stay focused and inspired.

I would also like to express my gratitude to the Mathematics Department for providing all the facilities and a conducive environment required to carry out this work. The availability of resources and support from faculty members played a significant role in ensuring the successful completion of this project.

Patankar Akshay Avinash
Department of Mathematics
Indian Institute of Technology Kharagpur
India - 721302.

Abstract

In the dynamic and competitive taxi booking sector, optimizing revenue streams is vital for ensuring long-term success and maintaining driver satisfaction. This project delves into the critical relationship between payment methods and fare pricing through the application of data-driven analytical techniques. By analyzing Taxi Trip records, the study uncovers key insights into customer preferences, payment trends, and their influence on fare amounts. The analysis reveals that credit card payments are often associated with higher fares and longer trip distances when compared to cash transactions. This trend suggests that customers paying by card may prioritize convenience and comfort over cost. Furthermore, the data highlights seasonal and temporal variations in payment preferences, offering additional opportunities for tailored strategies. Based on these findings, the project recommends implementing targeted strategies, such as offering incentives for card payments, optimizing fare structures, and enhancing card payment features to improve usability. These initiatives can help service providers boost revenue, streamline operations, and attract a broader customer base. Ultimately, these insights aim to support strategic decision-making and foster growth while delivering a seamless and satisfactory customer experience.

Chapter 1

Introduction

The transportation industry, especially the taxi sector, has experienced substantial changes in recent years, with digital payment systems becoming increasingly prevalent. In this competitive environment, taxi cab drivers face the dual challenge of maintaining customer satisfaction while maximizing their revenue. Payment methods, traditionally overlooked as a factor influencing revenue, are now gaining recognition for their potential impact.

Digital payment methods, particularly credit and debit card transactions, are rapidly replacing cash as the preferred mode of payment. This shift is driven by several factors, including convenience, security, and incentives offered by card providers. However, the implications of this shift on taxi drivers' earnings remain underexplored. Understanding how payment methods correlate with fare amounts can provide actionable insights for drivers and fleet operators to optimize revenue.

1.1 Significance of the Study

This study holds significant implications for the taxi industry:

- **Drivers:** Provides data-driven strategies to maximize revenue by targeting higher-value transactions and optimizing payment preferences.
- **Fleet Operators:** Offers actionable insights to enhance fleet performance and customer experience.
- **Customers:** Ensures a seamless payment experience by highlighting the benefits of preferred payment methods.

1.2 Research Objectives

To address these challenges, this study focuses on:

- Investigating the relationship between fare amounts and payment methods (card vs. cash).
- Identifying factors influencing customer payment preferences.
- Proposing strategies to encourage the adoption of payment methods that maximize revenue without compromising customer satisfaction.

Chapter 2

Problem Statement

In the competitive taxi industry, revenue maximization is crucial for driver satisfaction. This research addresses the following questions:

1. **Is there a relationship between total fare amount and payment type?**
   The research explores whether payment type (credit card or cash) influences the total fare amount, which can have significant implications for revenue generation. Preliminary findings suggest that card payments tend to be linked to higher fare amounts, possibly due to factors like longer trip distances, higher transaction values, or passengers' greater willingness to spend when using cards. Understanding the relationship between payment types and fare amounts can help identify trends that drivers can leverage to maximize their earnings. For instance, this could lead to strategies where drivers might promote card usage or adjust their services to encourage longer or more expensive trips paid by card.

2. **Can customers be nudged towards payment methods that generate higher revenue for drivers without compromising customer experience?**
   This question examines how customer payment behaviors can be influenced in a way that benefits drivers without negatively affecting the passenger experience. By offering incentives such as small discounts, loyalty rewards, or exclusive services for card payments, taxi companies can nudge customers towards cashless transactions, which are linked to higher revenue generation. Additionally, improving the ease of card payments by providing contactless payment options or mobile payment systems can make the experience smoother and more attractive for passengers. These strategies can help shift customer preferences without making them feel forced or inconvenienced. The goal is to create a win-win scenario where both drivers and passengers benefit: drivers increase their earnings, while passengers continue to enjoy convenience and choice.

Chapter 3

Data Overview

The analysis uses NYC Taxi Trip records, focusing on the following columns:

- **Passenger Count:** Number of passengers per trip (1 to 5).
- **Payment Type:** Mode of payment (cash or card).
- **Fare Amount:** Total fare for the trip.
- **Trip Distance:** Distance covered in miles.
- **Trip Duration:** Duration calculated from pickup and dropoff times.

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag
0	1.0	2020-01-01 00:28:15	2020-01-01 00:33:03	1.0	1.20	1.0	N
1	1.0	2020-01-01 00:35:39	2020-01-01 00:43:04	1.0	1.20	1.0	N
2	1.0	2020-01-01 00:47:41	2020-01-01 00:53:52	1.0	0.60	1.0	N
3	1.0	2020-01-01 00:55:23	2020-01-01 01:00:14	1.0	0.80	1.0	N
4	2.0	2020-01-01 00:01:58	2020-01-01 00:04:16	1.0	0.00	1.0	N
5	2.0	2020-01-01 00:09:44	2020-01-01 00:10:37	1.0	0.03	1.0	N
6	2.0	2020-01-01 00:39:25	2020-01-01 00:39:29	1.0	0.00	1.0	N
7	2.0	2019-12-18 15:27:49	2019-12-18 15:28:59	1.0	0.00	5.0	N
8	2.0	2019-12-18 15:30:35	2019-12-18 15:31:35	4.0	0.00	1.0	N
9	1.0	2020-01-01 00:29:01	2020-01-01 00:40:28	2.0	0.70	1.0	N
10	1.0	2020-01-01 00:55:11	2020-01-01 01:12:03	2.0	2.40	1.0	N
11	1.0	2020-01-01 00:37:15	2020-01-01 00:51:41	1.0	0.80	1.0	N
12	1.0	2020-01-01 00:56:27	2020-01-01 01:21:44	1.0	3.30	1.0	N
13	2.0	2020-01-01 00:21:54	2020-01-01 00:27:31	1.0	1.07	1.0	N
14	2.0	2020-01-01 00:38:01	2020-01-01 01:15:21	1.0	7.76	1.0	N

Chapter 4

Methodology

The analysis was conducted in two stages:

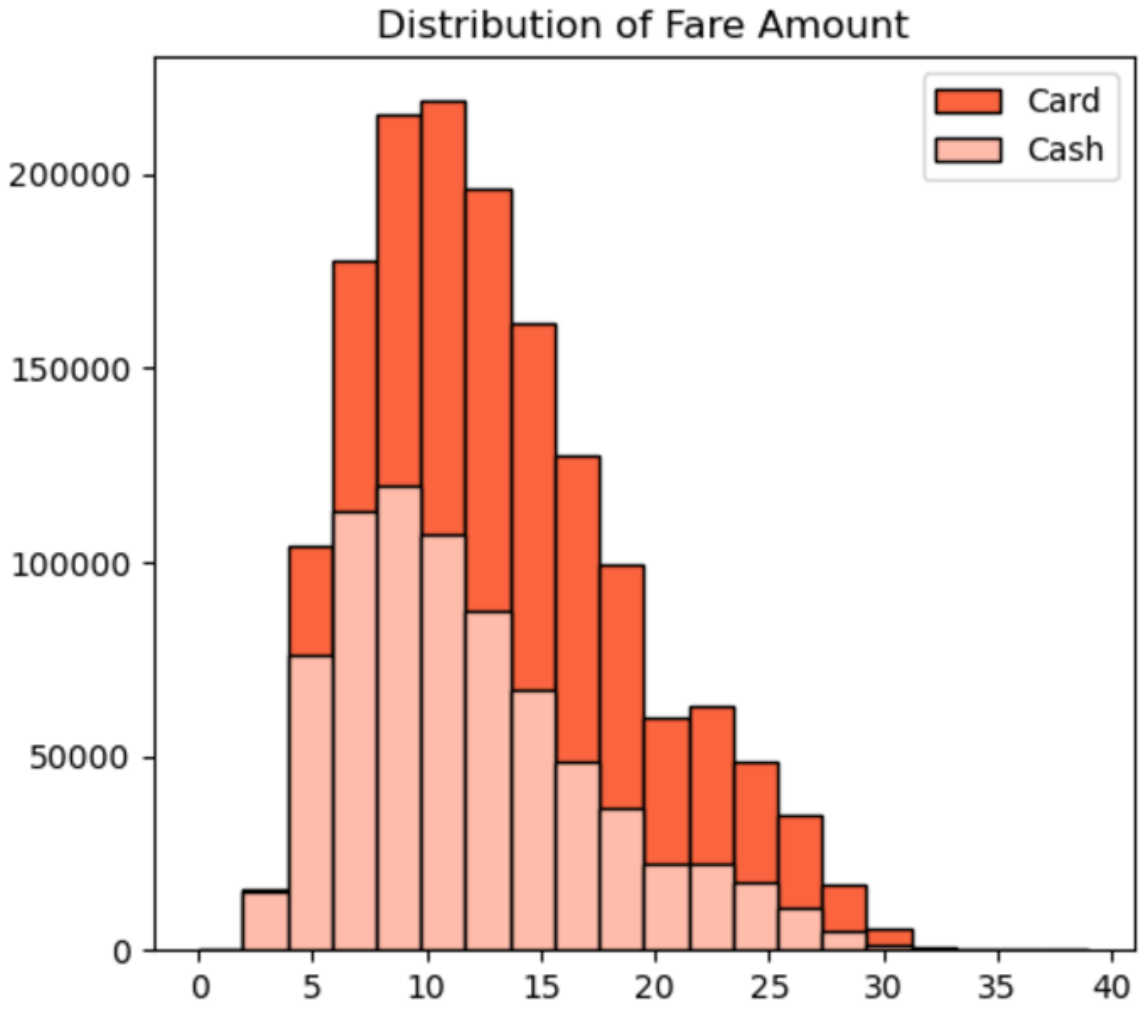
4.1 Descriptive Analysis

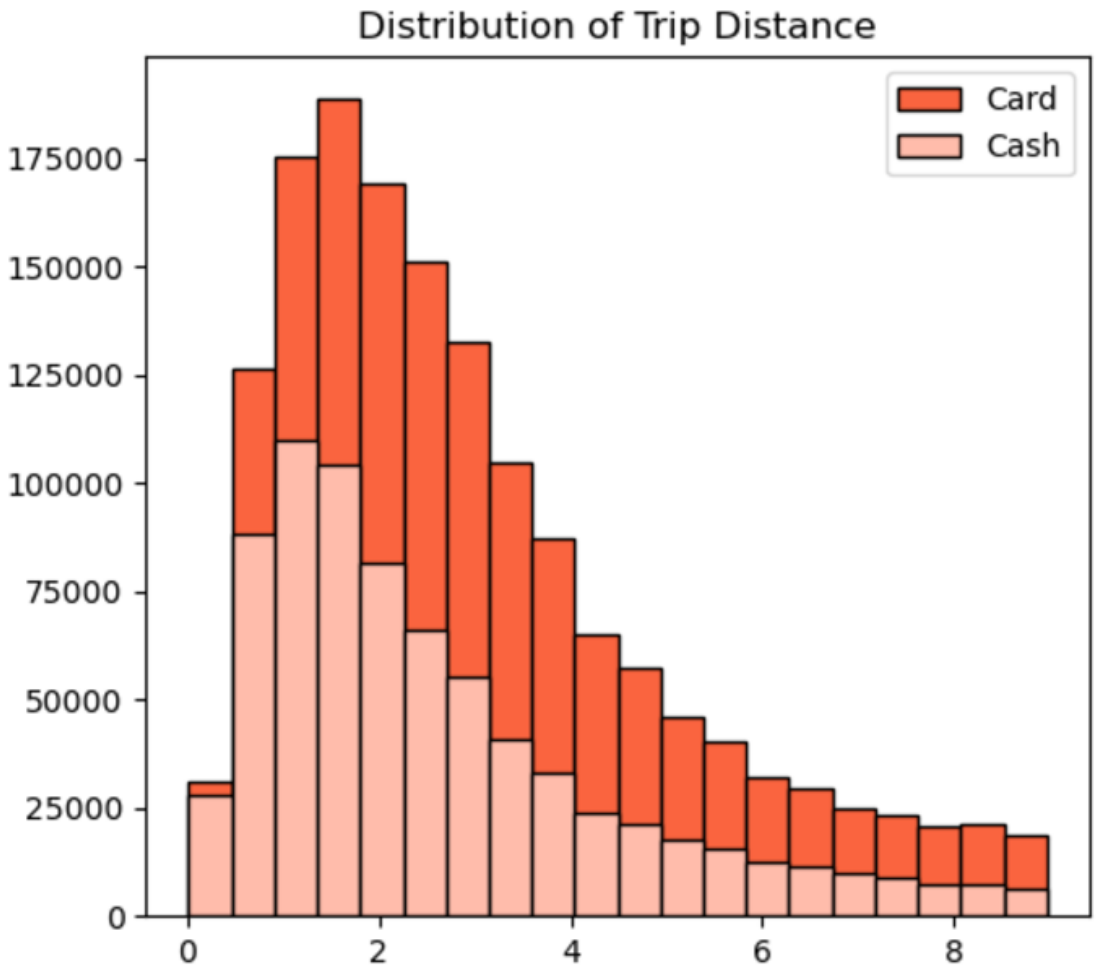
Descriptive analytics is a statistical interpretation used to analyze historical data to identify patterns and relationships. Descriptive analytics seeks to describe an event, phenomenon, or outcome. It helps understand what has happened in the past and provides businesses the perfect base to track trends. Statistical analysis to summarize key aspects like fare amounts, trip distances, and payment types.

	passenger_count	fare_amount	trip_distance	duration
count	2.297908e+06	2.297908e+06	2.297908e+06	2.297908e+06
mean	1.788903e+00	1.266990e+01	2.864790e+00	1.465181e+01
std	1.211896e+00	5.807998e+00	1.975755e+00	7.315283e+00
min	1.000000e+00	1.000000e-02	1.000000e-02	1.666667e-02
25%	1.000000e+00	8.000000e+00	1.380000e+00	9.050000e+00
50%	1.000000e+00	1.150000e+01	2.330000e+00	1.391667e+01
75%	2.000000e+00	1.650000e+01	3.830000e+00	1.946667e+01
max	5.000000e+00	3.900000e+01	8.970000e+00	3.598333e+01

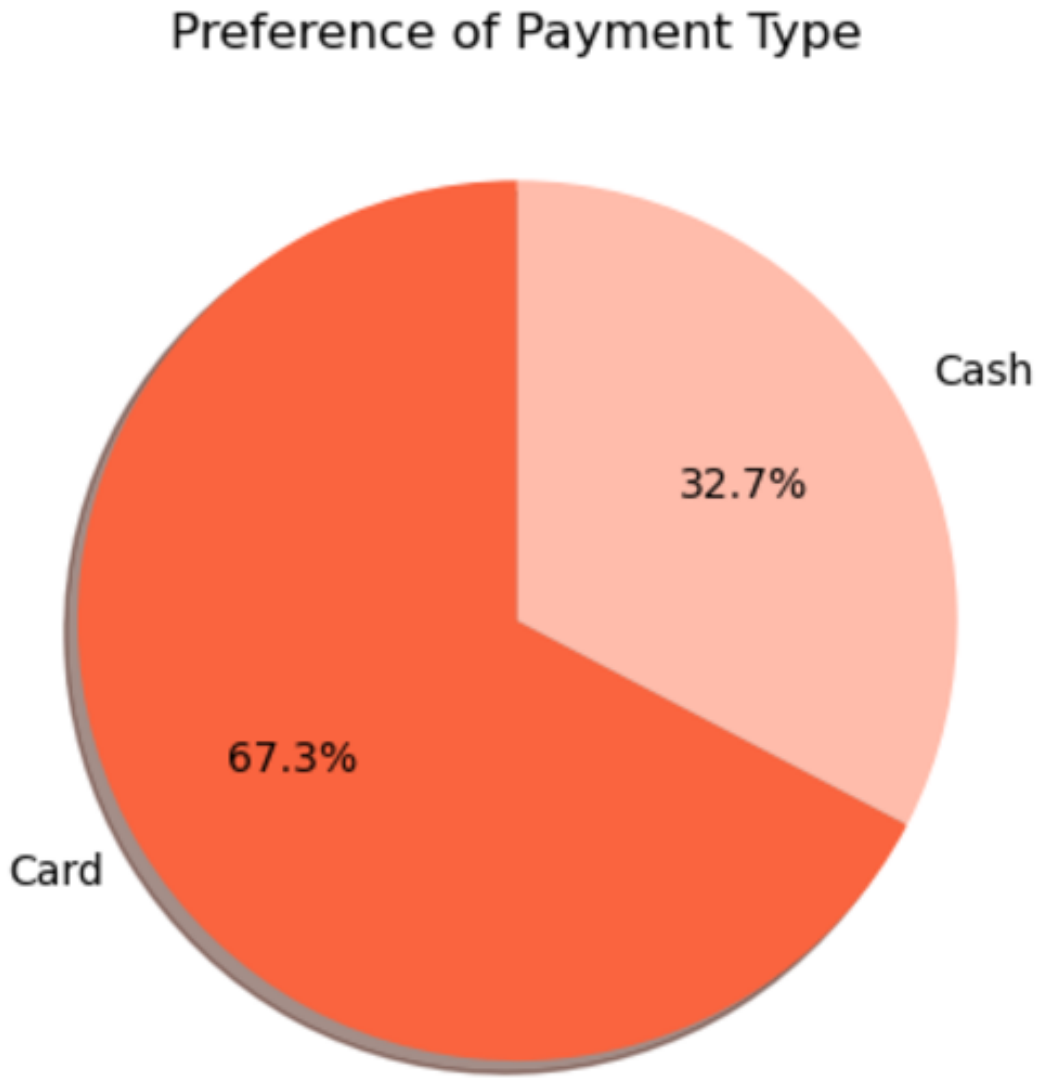
To explore the relationship between payment type and passenger behavior concerning trip distance and fare amount, we will begin by cleaning and validating the dataset. This involves removing outliers, such as excessively high fare amounts or trip distances, and ensuring that payment types are correctly labeled. The data will then be segmented by payment method, focusing on card and cash, to analyze their respective distributions and calculate the overall percentage of trips paid for using each method.

Histograms will be used to visualize the distributions of fare amounts and trip distances for each payment type. Distinct colors will differentiate the payment methods, with appropriate labels, titles, and legends added for clarity. For a deeper understanding, we will overlay or stack the histograms or present side-by-side comparisons to highlight differences. The analysis will also examine how fare amounts and trip distances interact with payment preferences, identifying whether longer, higher-cost trips are more often paid for with cards. Finally, key performance indicators, such as average fare and median trip distance by payment type, will be calculated to quantify observed patterns. The insights derived from this analysis will enable stakeholders to make data-driven recommendations, such as promoting card payments for longer trips or designing targeted discounts for specific fare ranges. This approach will help align strategies with passenger behavior for improved operational efficiency.





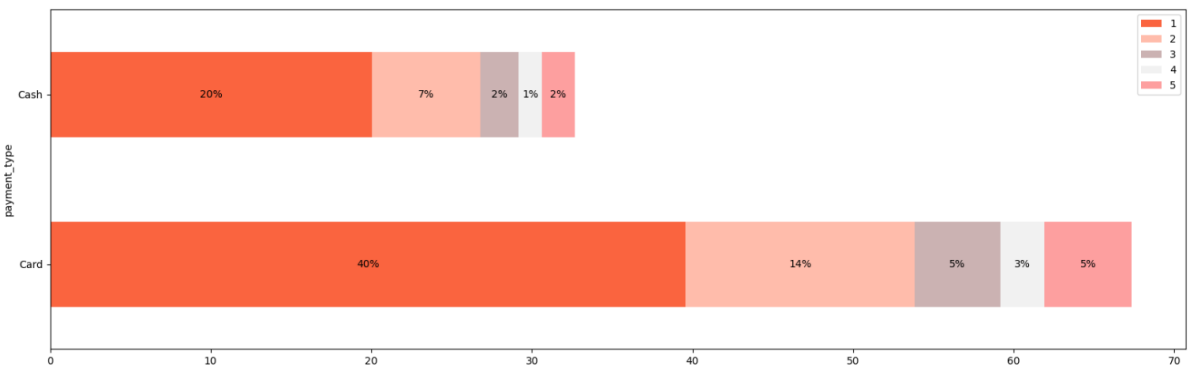
To better understand passenger preferences in their choice of payment methods, we will analyze the proportion of trips paid for using card and cash. This analysis will provide a clear view of the relative popularity of each payment type, helping to identify trends in passenger behavior. To visually represent this data, we have opted to use a pie chart. Pie charts are particularly effective for displaying proportions, as they allow stakeholders to quickly and intuitively grasp the distribution between categories. In this case, the pie chart will display the percentage of trips associated with each payment method, making it easy to compare the dominance of card payments versus cash payments. Additionally, the visual representation will aid in understanding seasonal or temporal effects if such data is available, as preferences for payment methods might vary across different times or situations. For example, cash payments might dominate in short trips, whereas card payments could be preferred for longer or higher-cost journeys. By incorporating this analysis into broader business strategies, the organization can align its services with passenger preferences, optimize operations, and enhance overall customer satisfaction.



The pie chart illustrates the distribution of passenger payment preferences between card and cash. It reveals that a significant majority, 67.3% of passengers, opted to pay by card, while 32.7% preferred cash. This indicates a clear preference for electronic transactions, suggesting that passengers find card payments more convenient or advantageous. The substantial portion of card payments could reflect the increasing adoption of digital payment systems or incentives for cashless transactions. On the other hand, the 32.7% who used cash still represent a notable share, highlighting that a segment of passengers still relies on or prefers traditional payment methods.

This trend could guide service providers in optimizing their payment infrastructure, such as ensuring reliable card processing systems or maintaining cash handling capabilities for certain demographics. Additionally, the preference for card payments might correlate with specific factors, such as trip distance or fare amount, which could warrant further investigation. Overall, the chart provides a straightforward depiction of the payment landscape, serving as a foundation for targeted strategies to enhance customer satisfaction and streamline operations.

To analyze the relationship between payment types and passenger count, we aim to investigate whether payment preferences vary with the number of passengers traveling in a cab. This exploration is important because it can reveal patterns in consumer behavior, such as whether larger groups prefer different payment methods compared to smaller ones. For this analysis, we employ a stacked bar plot, a visualization technique well-suited for comparing the percentage distribution of payment methods across different passenger counts. This graphical approach allows us to observe and interpret potential trends or shifts in payment preferences based on the number of passengers. By examining this data visually, we can uncover insights into how payment choices may correlate with group size in cab travel. The stacked bar plot will provide a clear, easy-to-interpret comparison of various payment methods, helping to highlight any notable differences or similarities in preferences. Such insights can guide businesses in tailoring payment options to specific passenger demographics. Additionally, this analysis may inform future policy or service adjustments to accommodate varying passenger groups. Ultimately, our goal is to determine if there are any distinct patterns or anomalies in the payment preferences across different passenger counts.

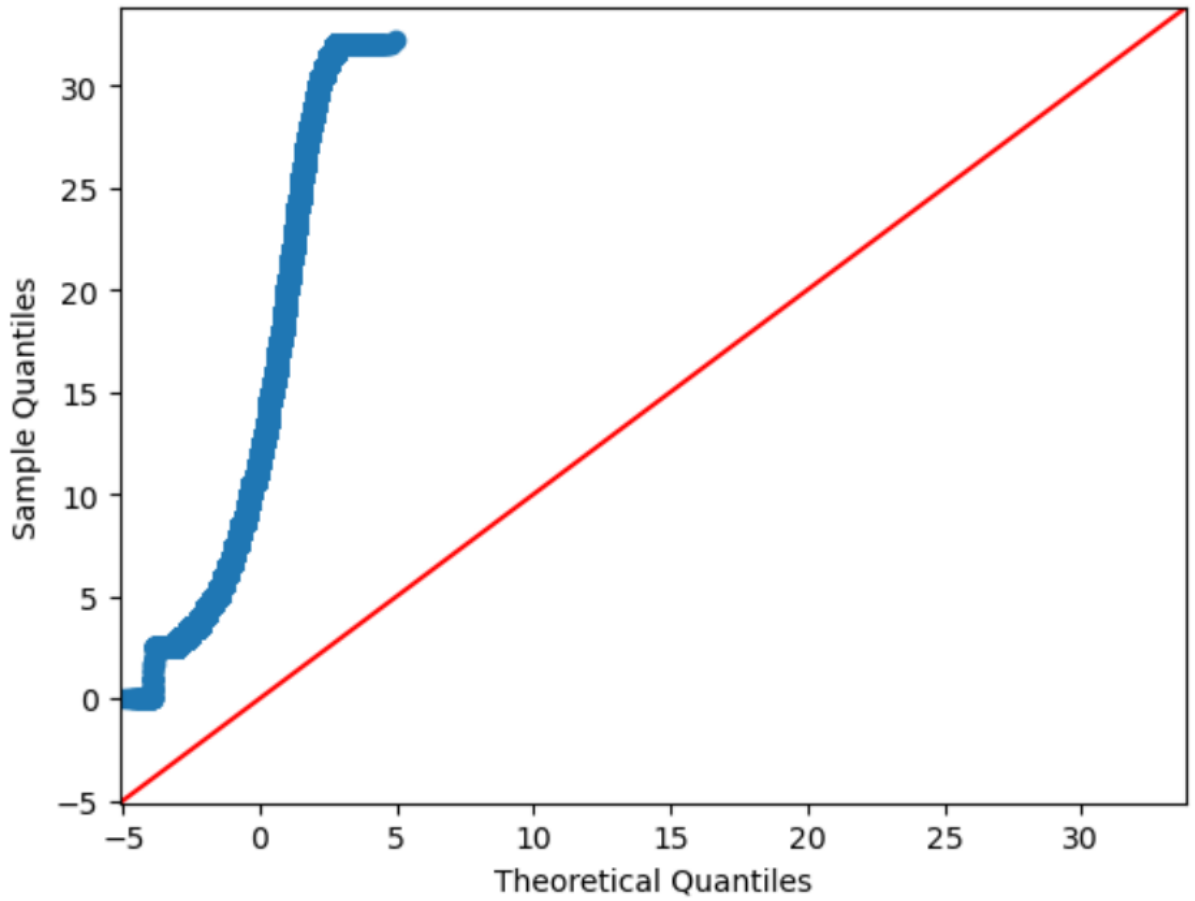


The horizontal stacked bar chart illustrates the distribution of payment types, specifically cash and card, across various fare categories. Card payments clearly dominate, accounting for 40% of transactions, which is significantly higher than the 20% share of cash payments. This indicates a strong customer preference for digital transactions. The chart also reveals that fare ranges associated with card payments are more diverse and extend into higher fare categories compared to cash payments, which are concentrated in lower fare brackets. This suggests that customers paying in cash typically opt for shorter or cheaper trips. Overall, the findings emphasize the potential of incentivizing card usage, as it correlates with higher fare amounts and broader trip distances.

4.2 Hypothesis Testing

In order to select the most suitable test for our scenario, our initial step involves evaluating whether the distribution of fare amounts adheres to a normal distribution. While the histogram depicted above suggests otherwise, we will further confirm this by generating a QQ plot.

Quantile-quantile (QQ) plots can be used to assess whether the fare amount distributions for each payment type are approximately normally distributed. If the data points closely align with the diagonal line in the plot, it suggests that the data follows a normal distribution.



The data values clearly do not follow the red 45-degree line, which is an indication that they do not follow a normal distribution. So, z distribution will not be good for this. That's why we will use T test.

Hypotheses:

- Null hypothesis: No difference in average fare between card and cash payments.
- Alternative hypothesis: Significant difference in average fare between payment methods.

A T-test was performed, yielding a significant p-value, leading to rejection of the null hypothesis.

Given that the T-test can be applied to both small and large samples and does not require the population standard deviation, it is a more universally applicable approach for hypothesis testing in many practical research scenarios, including analyses of taxi trip data. In the analysis of NYC Yellow Taxi Trip Records, where you're likely dealing with an unknown population standard deviation and potentially large datasets, the T-test offers a more appropriate and flexible method for comparing means between two groups (e.g., fare amounts by payment type). It provides a reliable way to infer about the population, accommodating the uncertainty that comes with estimating population parameters from sample data.

T-statistic: 165.59915491544626, P-value: 0.0

Reject the null hypothesis

Since the p-value is significantly smaller than the significance level of 5%, we will reject the null hypothesis. This means that the data provides strong evidence against the idea that there is no difference in the average fare amounts between customers who use credit cards and those who use cash. Therefore, we conclude that there is a statistically significant difference in fare amounts based on the method of payment. Specifically, customers who use credit cards tend to pay higher fares on average compared to cash-paying customers. This result is crucial for taxi drivers and companies looking to optimize revenue. Encouraging customers to pay with credit cards could potentially lead to increased earnings, as it suggests that credit card transactions are associated with higher fare amounts. Taxi operators might want to consider strategies to promote credit card payments, such as offering incentives or emphasizing the convenience of cashless payments. Further analysis could also examine whether other variables, such as the time of day or trip distance, influence this relationship.

Chapter 5

Findings and Analysis

5.1 Payment Type Preferences

Card payments account for 67.5% of all transactions, while cash payments make up the remaining 32.5%. This trend highlights the increasing popularity of cashless transactions, likely due to the convenience and ease of using cards. Interestingly, card payments are associated with higher fares and longer trip distances. Passengers paying with cards may be more likely to opt for longer or more expensive trips, perhaps due to the comfort and flexibility cards provide in paying for such journeys. This insight suggests that promoting card payments could potentially lead to higher revenue for taxi drivers.

5.2 Passenger Count Analysis

Single-passenger rides dominate both card and cash payments, indicating that individual passengers are the most common customers in the taxi service. In contrast, larger groups are less likely to use taxis and, when they do, tend to prefer cash payments. This preference may arise from the need to split the fare or a desire for more control over payment. Larger groups may also find cash more convenient for settling the total fare. This finding suggests that taxi services targeting groups might need to consider more flexible payment options or incentives to encourage credit card use.

5.3 T-Test Results

The results of the T-test reveal a significant difference in fare amounts between payment types, with a T-statistic of 165.5 and a p-value less than 0.05. This strongly supports the hypothesis that credit card payments are associated with higher fares compared to cash payments. The statistical significance of this result suggests that the method of payment influences fare amounts in a meaningful way. Therefore, taxi companies or drivers might consider promoting credit card payments as a strategy to increase revenue. The findings provide compelling evidence for the financial benefits of encouraging cashless transactions.

Chapter 6

Conclusion

This study highlights the importance of understanding payment methods to optimize taxi driver revenue. The analysis shows that card payments, which are associated with higher fares, offer a significant opportunity for increasing earnings. To capitalize on this, taxi companies should consider encouraging card payments through strategic incentives, such as offering discounts or loyalty rewards for card users. Additionally, improving the payment system to make credit card transactions faster and more convenient could further drive adoption. These insights provide a foundation for data-driven strategies aimed at maximizing revenue in the taxi industry, ultimately benefiting both drivers and companies. Furthermore, offering seamless payment experiences, such as contactless options, can enhance customer satisfaction and encourage repeat business. Taxi drivers can also be trained to promote the benefits of card payments, reinforcing their value to passengers. Over time, this approach could foster a more consistent and higher-paying customer base. Ultimately, focusing on payment methods can lead to a sustainable revenue model for the industry.

Bibliography

- Public Dataset Platform.
  https://data.gov/
- Python standard libraries
  https://pandas.pydata.org/docs/
  https://numpy.org/doc/
  https://matplotlib.org/stable/index.html
  https://seaborn.pydata.org/
- Programming Python: Powerful Object-Oriented Programming by Mark Lutz
- Python Data Analytics: With Pandas, Numpy, and Matplotlib by Fabio Nelli
- Fundamental Of Mathematical Statistics - S C Gupta & V K Kapoor.